Calibration and sensitivity analysis of long-term generation investment models using Bayesian emulation

M. Xu[a], A. Wilson[b], C.J. Dent[a]

[a]*School of Engineering and Computing Sciences, Durham University, DH1 3LE, UK*
[b]*Department of Mathematical Sciences, Durham University, DH1 3LE, UK*

Abstract

Investments in generation are high risk, and the introduction of renewable technologies exacerbated concern over capacity adequacy in future power systems. Long-term generation investment (LTGI) models are often used by policymakers to provide future projections given different input configurations. To understand both uncertainty around these projections and the ways they relate to the real-world, LTGI models can be calibrated and then used to make predictions or perform a sensitivity analysis (SA). However, LTGI models are generally computationally intensive and so only a limited number of simulations can be carried out. This paper demonstrates that the techniques of Bayesian emulation can be applied to efficiently perform calibration, prediction and SA for such complex LTGI models.

A case study relating to GB power system generation planning is presented. Calibration reduces the uncertainty over a subset of model inputs and estimates the discrepancy between the model and the real power system. A plausible range of future projections that is consistent with the available knowledge (both historical observations and expert knowledge) can be predicted. The most important uncertain inputs are identified through a comprehensive SA. The results show that the use of calibration and SA approaches enables better decision making for both investors and policymakers.

Keywords: Generation investments, calibration, uncertainty analysis, sensitivity analysis, Bayesian emulation

Nomenclature

Sets and Functions

$\mathcal{T}$	Set of planning years of interest, indexed by t.
$\mathcal{P}$	Set of past planning years.

Email addresses: meng.xu4@durham.ac.uk (M. Xu), amy.wilson@durham.ac.uk (A. Wilson), chris.dent@durham.ac.uk (C.J. Dent)

$\mathcal{F}$	Set of future planning years.
G	Set of generation technologies, indexed by g.
J	Subset of input variables, indexed by i, j.
ρ	Set of fuel types including uranium, coal, gas, and carbon.
$f(\cdot), \tilde{f}(\cdot)$	Functions of the simulator and the emulator, respectively.
$h_i(\cdot)$	Functions of the main modules within the simulator.
$\mathcal{GP}(\cdot, \cdot)$	Gaussian process function.
$p(\cdot)$	Probability distribution function.

Parameters and Variables

x	Vector of input variables.
u, θ, ω	Vector of control inputs, calibration parameters and forcing inputs, respectively.
I	Total number of input variables.
$y^B_{g,t}, y^M_{g,t}, y^D_{g,t}$	Investment, mothballing and de-mothballing of generation capacities of type g at year t, respectively.
$y_{g,t}$	Installed generation capacity of type g in operation at year t.
y_{obs}	Vector of historical observations of elements $y_{obs,t}$ over $\mathcal{P}$.
$F_{\rho,t}$	Fuel price of type ρ at year t.
MR_ρ	Reference trend level of annual fuel prices of type ρ.
P_ρ	Multiplier applied to the trend level of fuel type ρ.
$P_{markup,t}$	Hourly price markup payment at year t.
θ_{markup}	Markup cut-in point where the markup approaches to zero.
ND_t	Hourly net demand (demand minus wind generation) at year t.
$AG_{g,t}$	Hourly available thermal capacity subject to forced outages at year t.
CM_t	Hourly capacity margin (hourly available thermal capacity minus hourly net demand) at year t.
$VOLL$	Value of lost load (VOLL).
$P_{e,t}$	Energy prices at year t.
$C_{g,t}$	Generation cost of generation type g at year t.
$LOLE_t$	Loss-of-load expectation at year t.
$RT_{g,t}$	Retirement of existing generators of type g at year t.
V_t	Net Present Value (NPV) of an investment at year t.
τ_f	The furthest simulation year ahead of the current decision year.
$V_{VaR,t}$	Value at Risk (VaR) of V_t.
θ_{VaR}	Assumed level of risk aversion.
β, σ^2, γ	Hyperparameters in the Gaussian Process model.
δ	Model discrepancy function.
D	Design points of chosen input variables.
K_f	Principal component basis vectors of elements $k_1, \cdots, k_{p_f}$.
d_i	The i-th basis function for model discrepancy.
ϑ_i	Weight of the i-th basis function for model discrepancy.
λ_ϑ	Hyperparameter in the model discrepancy.
p_δ	Total number of basis functions for model discrepancy.

S_{J}	Measure of sensitivity to a subset of inputs x_{J}.
SV_{J}	Variance of the main effect of a subset of inputs x_{J}.
$\mathrm{Var}(y)$	Total output variance.

1. Introduction

There is a growing concern over capacity adequacy in future power systems due to a number of risks that may discourage investment in generation capacity [1–4]. These risks exposed to investors range from policy (e.g., VOLL pricing, CO_2 prices and renewable targets) and market (e.g., fuel cost, demand forecast and electricity price) risks, to technology (e.g., capital cost) and finance (e.g., hurdle rate) risks [4] and they create uncertainty (*i.e.*, imperfect knowledge) in the financial returns of an investment. One prominent feature in future power systems is that market risks increase with the amount of variable wind power that contributes to higher price volatility and lower (on average) and more uncertain load factors for thermal power plants [5].

Various long-term generation investment (LTGI) models have been developed for predicting real-world generation projections and hence guiding investment decisions and the design of energy policy [6–13]. From the perspective of policymakers, who wish to adequately account for uncertainty around future generation projections related to the real world, it becomes increasingly important to consider two main sources of uncertainty existing in these models. One is input uncertainty representing investment risks and/or model assumptions that affect or shape the direction of investment decisions [4]. The other one is structural uncertainty which concerns the discrepancy between the model and the real-world complex investment decision-making process. Questions regarding validation and understanding of these LTGI models need to be carefully addressed before model outcomes can be interpreted and applied.

Calibration or *history matching* is a valuable tool for validating a model and linking it to the real world when historical observations are available. This typically involves calibration of a subset of uncertain model parameters against historical observations of the model output whilst modeling the discrepancy between the model and the real system. Uncertainty of calibration parameters, which may be specified ex ante as a probability distribution based on the prior beliefs of the model user or other experts, can be reduced through calibration (*i.e.*, by identifying values of calibration parameters that are plausible with respect to prior beliefs and historical observations of the model output). To the best of our knowledge, no such formal calibration of LTGI models has previously been done. If a calibration against historical observations is not performed, this severely limits the conclusions which can be drawn regarding investment decisions and policy design in the real system.

Sensitivity analysis (SA) is also often applied to LTGI models in order to understand how model outputs react to changes in model inputs. SA in [6, 9, 11, 13–15] were carried out using a simple one-at-a-time method, where each uncertain parameter is varied independently across a range of possible values while all others are held constant. The one-at-a-time method fails to treat the

analysis with sufficient care (*i.e.*, no formal weight or probability is attached to each outcome), and is incapable of taking into account interactions among different inputs. Multi-way SA can identify the combined effects of two or more inputs, through varying the inputs together using a large and highly structured set of simulator runs [16]. Probabilistic SA is an alternative approach to multi-way SA that can address interactions and nonlinearities. The input uncertainty is explicitly described as a scenario tree with associated probabilities (discrete) or a probability distribution (continuous) in probabilistic SA while it is treated only implicitly in the preceding methods. A wide ranging review of uncertainty and sensitivity analysis in the context of power system planning may be found in [17].

A conventional way to conduct a formal calibration or a probabilistic SA is the Monte Carlo (MC) method of drawing random configurations of inputs from their uncertainty distributions, running the model for each input configuration to obtain the set of outputs, and constructing the output distribution (which can in principle be evaluated to any desired accuracy). Computationally intensive models associated with large studies tend to have high-dimensional inputs. The MC-based method may require thousands of (if not more) individual evaluations in order to avoid sparse coverage of the model input space. It may be practically impossible for complex models to achieve very dense coverage of input space even if very large computer resource is available [17]. For example, a single run of a LTGI model may take many hours [6, 18] or even many days or weeks with more detailed modeling of short-term operations of power plants [19, 20]. In addition, the outputs of interest (e.g., generation projections) for a LTGI model are often high-dimensional due to the long planning horizon; this adds to the complexity of calibration and SA. Even where a very large number of runs may be possible by acquiring additional computing resource, the approach adopted in this paper allows results to be obtained in a systematic way with a smaller computing resource.

This paper will carry out calibration and probabilistic SA of a computationally intensive LTGI model (*i.e.*, *the simulator*) with careful management of two sources of uncertainty - input uncertainty and structural uncertainty. A highly-efficient Bayesian approach described in [21–24] is employed. Fig. 1 shows a diagram of the proposed Bayesian framework, which is based on a Gaussian process model (*i.e.*, *the emulator*) that is built as an approximation of the simulator using a limited number of simulation runs (*i.e.*, *training data*). The emulator can efficiently deal with the tasks of: *calibration*; *probabilistic SA*; *prediction* - estimation of model outputs at input configurations that have not been tested; and *uncertainty analysis* that is most relevant when those outputs provide guidance in the making of some decision (such as using a LTGI simulator in setting VOLL for maintaining the LOLE target).

The main contributions of this paper can be summarized as follows.

1) Use of Bayesian emulation to manage uncertainties arising from the limited number of runs that are possible and consequent sparse coverage of the input space; this is the first time that such emulation techniques have been

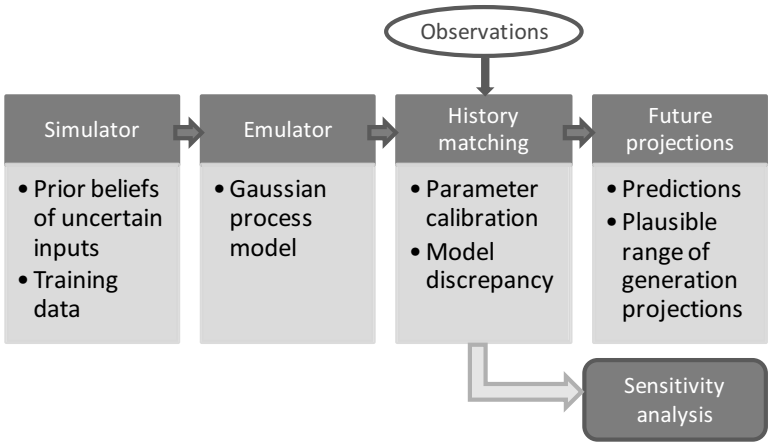

Figure 1: Steps involved in performing the Bayesian framework based on emulation.

used to manage these uncertainties when performing model calibration and uncertainty analysis associated with generation investment.

2) Presenting a statistical approach for the calibration of LTGI models, to infer from historical observations improved knowledge of uncertain inputs and model discrepancy;

3) Quantifying a plausible range of model outputs that is consistent with the available knowledge (both historical observations and expert knowledge), and demonstrating that a failure to account for the parameter and the structural uncertainty may mean that results are misleading to investors and policymakers;

4) Performing an efficient SA of a LTGI model and identifying the most important uncertain model inputs.

In Section 2, a brief description of the simulator under study is provided, emphasizing uncertain model inputs and outputs that are of interest. The theoretical foundations of the Bayesian framework is illustrated in Section 3. This Bayesian approach is applied to the GB power system as a case study in Section 4, including results from calibration, predictions and SA as well as discussions on the generality of the Bayesian approach. Some conclusions are drawn in Section 5.

2. The Long-term Generation Investment Simulator

A brief description of the simulator under study is provided in this section, with emphasis on the uncertain model inputs and the output of interest. A full description of the simulator may be found in [6, 18]. This LTGI model developed by Eager, Bialek and Hobbs [6, 18] will be used as an exemplar to demonstrate

the application of the Bayesian approach to a decision-support tool. Whilst this is a specific application, the Bayesian framework as presented here can be generally applied to models in which uncertainty plays a key role and where the link between the model and reality is of great importance and interest to model users. For more discussions, see Section 4.6.

2.1. High-level formulation of the generation investment model

The simulator can be used to quantify thermal investments given a scenario of on-shore and off-shore wind capacities over a planning horizon $\mathcal{T}$. The investment logic is based on an NPV approach combined with a VaR criterion applied to the assessment of the profitability of investments based on forward-looking simulations. A MC approach is employed to account for uncertain demand growth and fuel and carbon prices at each simulation year, and estimates the probability distribution of profitability at the decision year. Fig. 2 shows the structure and the main inputs and outputs of the simulator. A high level description of the five main modules within the simulator is provided in eqs. (1a)–(1e).

$$F_{\rho,t} = h_1\big(P_\rho * MR_\rho\big), \quad \forall \rho \in \{uranium, coal, gas, carbon\} \tag{1a}$$

$$P_{markup,t} = h_2\big(ND_t, \{AG_{g,t}\}, \theta_{markup}, VOLL\big), \quad \forall g \in \mathrm{G} \tag{1b}$$

$$\big[P_{e,t}, \{C_{g,t}\}, LOLE_t\big] = h_3\big(ND_t, \{AG_{g,t}\}, F_{\rho,t}, P_{markup,t}\big), \quad \forall g \in \mathrm{G} \tag{1c}$$

$$\big[y^B_{g,t}, y^M_{g,t}, y^D_{g,t}\big] = h_4\big(\theta_{VaR}, \{P_{e,\tau}\}, \{C_{g,\tau}\}\big), \quad \forall \tau \in \{t, \cdots, t+\tau_f-1\} \tag{1d}$$

$$y_{g,t} = h_5\big(y_{g,t-1}, y^B_{g,t}, y^M_{g,t}, y^D_{g,t}, RT_{g,t}\big), \tag{1e}$$

In (1a), $h_1(\cdot)$ simulates annual fuel and carbon prices (both in the past and future) using mean-reverting stochastic processes [6, 25] which address volatility around the trend (*i.e.*, mean-reversion) levels of fuel prices. A multiplier P_ρ adjusts the reference trend level MR_ρ upwards or downwards, to reflect the long-run or global uncertainty of fuel prices due to market changes or political interventions. Applying a multiplier P_{carbon} to the trend of future carbon prices, as shown in Fig. 3, results in different levels of carbon price trend (in solid and dashed lines) associated with modeled local uncertainty (in shaded areas).

In (1b), $h_2(\cdot)$ calculates the price markup (*i.e*, an uplift), referring to charges applied in addition to a uniform market-clearing price during periods of tight supply (*i.e.*, scarcity situations), as shown in the upper graph in Fig. 4. The price markup, $P_{markup,t} := VOLL * e^{b*CM_t}$, is represented as an exponential function of the capacity margin. The markup reaches VOLL at a capacity margin of zero, since VOLL is used as the price cap. The parameter b is calibrated so that the markup approaches zero at the value of θ_{markup}. Different assumptions on θ_{markup} lead to different levels of price markup, as shown in the lower graph in Fig. 4.

In (1c), $h_3(\cdot)$ performs the probabilistic production costing method in a nonequilibrium market settlement, as described in [6]. The method performs a convolution of generator outages with the annual net demand curve ND_t and calculates the energy prices, costs, revenues and LOLE. ND_t takes into account the relationships between wind availability and demand and is modeled by a

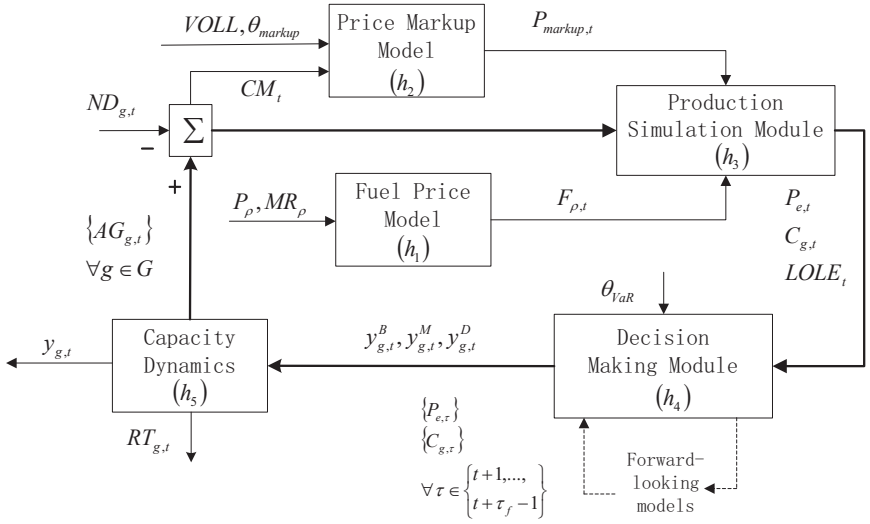

Figure 2: The model structure of the simulator.

mix of normal distributions. Annual load growth rates in forward-looking simulations are sampled from independent normal distributions by a MC approach. Since an existing published model is adopted, discussion about more details on the form of this model is beyond the scope of the paper.

In (1d), $h_4(\cdot)$ determines the investment in, mothballing and de-mothballing of existing thermal generators. At each year, the investor assesses V_t based on the NPV of the first τ_f years of forecasted profits. Note that V_t is random due to the uncertain nature of fuel prices, demand growth rates and generator outages. The VaR criterion is applied to V_t according to $P_r(V_t \leq V_{VaR,t}) = \theta_{VaR}$. The smaller the value of θ_{VaR} is, the more risk averse the investor is assumed, and the lower level of investment would be.

The last function $h_5(\cdot)$ in (1e) simulates capacity dynamics after taking into the decisions derived from (1d) and exogenous thermal plant retirements $RT_{g,t}$.

2.2. Model inputs and outputs

The inputs of interest here are sources of uncertainty that can result in a substantially different trend in the long-run investment decisions. Among the modules described in Section 2.1, five model inputs are identified and categorized into three types: control variables $u := \{VOLL, P_{carbon}\}$ that are determined by the VOLL pricing policy and the carbon policy, respectively; calibration parameters $\theta := \{\theta_{markup}, \theta_{VaR}\}$ with unknown values that will be learned using historical observations of the model output [21], as further explained in Section 3.3; and a forcing input $\omega := P_{gas}$ reflecting uncertainty in future gas prices. The parameters $VOLL$, P_{gas}, and P_{carbon} are not chosen to be calibrated since they are affected by market changes or political interventions, whereas calibration parameters must take the same value for both historical

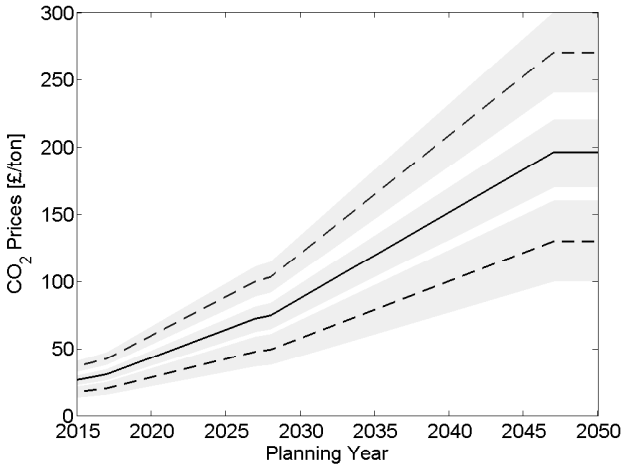

Figure 3: Different trend levels of carbon prices (in solid and dashed lines) using the central estimates of carbon prices published by DECC [26] as a reference (in solid line); the expanding shaded area reflects increasing local uncertainty further out along the trend level.

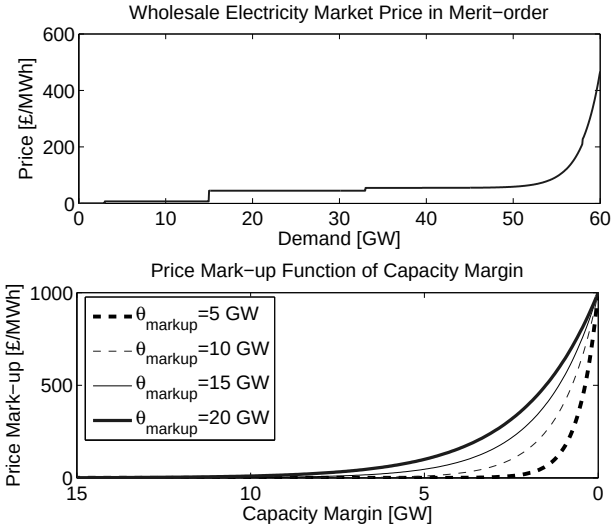

Figure 4: Upper: Energy prices cleared in merit-order with price markup; Lower: Price markup functions with different values of θ_{markup}.

observations and future projections. The above input categorization is derived from a policymaker's perspective; an investor might have a different categorization of inputs. The selected model inputs are assumed to be constant over the planning horizon. This approach allows direct exploration of the relationship between one parameter (used to represent a model input) and model outputs. Similar practices can be found in [9, 11, 13, 15]. The dimensionality of the input space can be extended if independent values are needed for an input variable at each year or at each stage (e.g., every five years). Also more model parameters (e.g., the mean or the variability of demand growth rates) can be included in the input space if they are deemed to contribute substantial additional uncertainty.

The output of interest is a time series of annual installed thermal generation capacity $y = \{y_t\}, \forall t \in 1 \dots \mathcal{T}$, where $y_t = \sum_{g \in \mathrm{G}} y_{g,t}$. The historical observations of installed thermal capacities are available, which allows for calibration in the history matching procedure. The planning time horizon of interest $\mathcal{T}$ is either the past ($t \in \mathcal{P}$), for which observations exist, or the future ($t \in \mathcal{F}$), for which a projection is made.

Given a set of model inputs $x = \{u, \theta, \omega\}$ of I elements, the simulator combines the five functions defined in eqs. (1a)–(1d) to give the output y. In this way, the simulator can be thought of as a deterministic function, f, with

$$y = f(x). \tag{2}$$

3. Bayesian framework based on emulation

In the LTGI model presented in Section 2.1, there is uncertainty in the input values of x, which propagates into output uncertainty, resulting in a range of generation projections. LTGI simulators are often too expensive to evaluate at a large number of points within the input space [18, 19, 27] (see Section 4.2 for details), which makes it difficult to identify the plausible range of generation projections that is consistent with historical observations.

The aim of Bayesian emulation is to evaluate the function f given in (2) at a small number of carefully configured input points, and to approximate this function as accurately as possible with a statistical representation $\tilde{f}$. The statistical representation should be computationally less demanding to evaluate than f, and will include terms modeling the uncertainty in the approximation at any point x in the input space where $f(x)$ has not been evaluated. The challenge posed by the high-dimensional model output in the emulation process is handled by a dimension reduction technique. Once the Bayesian emulator is built in the first-stage, problems such as prediction or SA can be efficiently tackled using the emulator in the second stage. The basic theory is provided in the following subsections. More detailed theoretical foundations can be found in [21, 22].

3.1. Emulation using a Gaussian process

The simulator represented by the function f in (2) is treated as an *uncertain function*, as the value of $f(x)$ for any value of x is unknown until the simulator is

run at x. Our prior uncertainty on the function f is modelled with a Gaussian process (GP) $\tilde{f}(\cdot) = \mathcal{GP}(\cdot,\cdot)$, with mean function $\mathrm{E}[f(x) \mid \beta]$ and covariance function $\mathrm{Cov}[f(x), f(x') \mid \sigma^2, \gamma]$ [21], so that,

$$f(x)|\beta,\sigma^2,\gamma \sim \tilde{f}(x) = \mathcal{GP}\Big(\mathrm{E}[f(x) \mid \beta], \mathrm{Cov}[f(x), f(x') \mid \sigma^2, \gamma]\Big), \quad x, x' \in [0,1]^I, \tag{3}$$

where

$$\begin{aligned} \mathrm{E}[f(x)|\beta] &= x^T\beta, \\ \mathrm{Cov}[f(x), f(x')|\sigma^2,\gamma] &= \sigma^2 exp\{-(x-x')^T\gamma(x-x')\}. \end{aligned} \tag{4}$$

The mean function depends on a vector of hyperparameters (*i.e.*,uncertain parameters) β and the covariance function is conditional on a hyperparameter σ^2 and a diagonal matrix γ of correlation hyperparameters; x, x' are any two points over the standardized input space $[0,1]^I$.

The forms of the prior mean $\mathrm{E}[f(x) \mid \beta]$ and covariance function $\mathrm{Cov}[f(x), f(x') \mid \sigma^2, \gamma]$ in the GP used here are given by,

$$\begin{aligned} \mathrm{E}[f(x) \mid \beta] &= 0, \\ \mathrm{Cov}[f(x), f(x') \mid \sigma^2, \gamma] &= exp[-\sum_{i=1}^{I}\{(x_i - x_i')/0.3\}^2], \end{aligned} \tag{5}$$

where the index i denotes the i-th element of the input. The hyperparameters β are set to zero and $\{\sigma^2, \gamma\}$ in (4) are fixed at specific values.

Given the prior GP model, the following three steps are taken to develop the emulator:

Step 1: Defining the standardized input space of interest through prior knowledge of these parameters; and selecting a small set of well designed input configurations, known as design points, $D := [x^{(1)}, x^{(2)}, \ldots, x^{(d)}]$ of d elements using latin hypercube designs [28].

Step 2: Running the simulator $f(\cdot)$ at each of these design points, and obtaining the simulator output $f(D) := (f(x^{(1)}), f(x^{(2)}), \ldots, f(x^{(d)}))^T$ of d elements.

Step 3: Fitting an emulator by combining the training data $(D, f(D))$ with the prior GP model given in (3). This procedure uses a Bayesian approach described in [24];

The trained emulator $\tilde{f}$ is a statistical distribution. In particular, for any point x', the distribution of the output $[f(x')|D, f(D), \beta, \sigma^2, \gamma]$ is another GP (for more details, including details on the estimation of $\{\beta, \sigma^2, \gamma\}$, see [29]). The emulator estimates the model response at any point in the parameter space and quantifies the uncertainty in this estimate; this means that computational tasks such as calibration, prediction and SA can be carried out efficiently. Validation is required to ensure that an emulator is sufficiently accurate. A number of diagnostics are provided in [30], such as the Mahalanobis distance, the analysis of prediction errors (see Section 4.3.1), the pivoted Cholesky decomposition etc.

3.2. Dimension Reduction

As described in Section 2.2, the simulator output of interest $y := f(x)$ is a vector of $\mathcal{T}$ elements. To cope with the high-dimensional model output space, principal components analysis (PCA), described in [31] is used to project the high dimensional output data into a new lower dimensional representation of the data that contains most of the variance in the data with minimal loss of information. The principal component basis vectors $K_f = [k_1, \cdots, k_{p_f}]$ are obtained via singular value decomposition of the standardized simulation output matrix $f(D)$. Based on PCA, the $\mathcal{T}$-dimensional simulator output $f(x)$ is modelled using a p_f-dimensional basis representation [24]:

$$f(x) \sim \sum_{i=1}^{p_f} k_i \nu_i(x), \tag{6}$$

where $\nu_i(x)$ are GP models described in (3). With this formulation, the problem of building an emulator that maps $[0,1]^I$ to $\mathbb{R}^{\mathcal{T}}$ is reduced to building p_f independent, univariate GP models for each $\nu_i(x)$.

3.3. Bayesian Calibration

The goal of model calibration here is to identify plausible values of calibration parameters whilst simultaneously inferring the model discrepancy using physical observations of the output over the time period $\mathcal{P}$. Among the inputs $x := \{u, \theta, \omega\}$ described in Section 2.2, $\{u, \omega\}$ are already known historically. The calibration parameters θ are assumed to have unknown best values due to our incomplete knowledge of the real-world. If the simulator were run with these best values, it would reproduce the observations plus a model discrepancy term δ (if we assume no observation error). With the emulator $\tilde{f}(\cdot)$ as an approximation of the simulator, the relationship between the observations, the model discrepancy and the emulator at the best value of θ can be written as [32],

$$y_{obs} = \tilde{f}(u, \theta, \omega) + \delta, \tag{7}$$

where $y_{obs} := \{y_{obs,1}, \cdots, y_{obs,\mathcal{P}}\}$ is the observation.

The model discrepancy δ quantifies the mismatch between the model and the observations at the best setting for the calibration parameter θ. The mismatch may arise from inadequacies in the simulator, such as in the model equations, model structure or logic [32]. δ is modelled by a linear combination of basis functions [33]:

$$\delta = \sum_{i=1}^{p_\delta} d_i \vartheta_i, \tag{8}$$

where d_i's are basis functions; and the weights ϑ_i's are modelled as iid $\mathrm{N}(0, \lambda_\vartheta)$ where the prior of λ_ϑ can be suggested by expert knowledge (*i.e.*, the knowledge about how accurately the simulator represents the system). Here the basis functions are independent normal kernels of which the number and the width depend on the application (see Section 4.3). More details regarding the use of

kernels to model the discrepancy against real-world observations can be found in [33].

To find plausible values for the calibration parameters, θ, alongside inferring the model discrepancy, first suppose that the prior knowledge of the calibration parameters, the discrepancy parameters λ_ϑ and the emulator parameters β, σ^2 and γ is described by the joint prior distribution $p(\theta, \lambda_\vartheta, \beta, \sigma^2, \gamma)$. This prior distribution is updated using the observations and the set of training runs obtained from the simulator. This updating is done according to Bayes' rule [32]:

$$p(\theta, \lambda_\vartheta, \beta, \sigma^2, \gamma | y_{obs}, f(D)) \propto p(y_{obs}, f(D) | \theta, \lambda_\vartheta, \beta, \sigma^2, \gamma) p(\theta, \lambda_\vartheta, \beta, \sigma^2, \gamma), \quad (9)$$

where the left hand term of 9 is the posterior distribution of the key parameters of interest and $p(y_{obs}, f(D)|\theta, \lambda_\vartheta, \beta, \sigma^2, \gamma)$ is the joint distribution of the observed data and the training runs $f(D)$, conditional on these key parameters. The marginal posterior distribution for each of $\theta, \lambda_\vartheta, \beta, \sigma^2$ and γ can be obtained through marginalisation over the other parameters. For discussion and technical details regarding this Bayesian calibration approach see [32].

3.4. Sensitivity analysis

Sensitivity analysis is used to quantify how much of the total output uncertainty is attributed to uncertainty in a particular input or a group of inputs. Using the calibrated emulator, a probabilistic SA that treats uncertainty explicitly may be carried out. Compared with that conducted in [17] where uncertainties are modeled as decision trees, the probabilistic sensitivity study formally treats uncertainties (inputs x, the emulator $\tilde{f}$, and outputs y) as probability distributions, as is required in the full Bayesian approach adopted in emulation.

Uncertainties are prioritized through a direct variance-based measure of sensitivity, known as "first-order effect index" or "main effect index" [16]. The sensitivity index S_J for inputs x_J is defined as the ratio of the sensitivity variance to the overall variance:

$$S_\mathrm{J} = \frac{SV_\mathrm{J}}{\mathrm{Var}(y)}, \quad (10)$$

where the total variance can be partitioned into the sum of all the sensitivity and interaction variances,

$$\mathrm{Var}(y) = \sum_{i=1}^{I} SV_i + \sum_{i<j}^{I} SV_{ij} + \cdots + SV_{12\ldots I}, \quad (11)$$

$$SV_i = \mathrm{Var}(E(\tilde{f}(x)|x_i)) = \mathrm{Var}(\int \tilde{f}(x) p_{-i|i}(x_{-i}|x_i) dx_{-i}), \quad (12)$$

$$SV_{ij} = \mathrm{Var}(E(\tilde{f}(x)|x_{ij}))) = \mathrm{Var}(\int \tilde{f}(x) p_{-ij|ij}(x_{-ij}|x_{ij}) dx_{-ij}), \quad (13)$$

where $E(\tilde{f}(x)|x_i)$ denotes expected value of the emulator at x, conditional on x_i, averaged over the joint distribution of all the other input variables x_{-i} (known as the mean effect); $E(\tilde{f}(x)|x_{ij})$ represents the two-input mean effect.

The sensitivity index can be quantified through exploring the distributions of model inputs together with the emulator via Markov chain Monte Carlo [23, 34].

4. GB case study

In this section, as a detailed case study, Bayesian emulation is applied to the GB power system with an energy-only market design. In particular, the simulator will be calibrated, as described in Section 3.3, quantifying the plausible range of generation projections, and a probabilistic SA will be performed. Note that the Bayesian framework presented in this paper is applicable to other types of generation investment model.

4.1. Data

The data used in the case study are consistent with those provided in [18], including the initial capacity mix, wind and demand data, as well as financial and technical assumptions for generators.

Expert knowledge can be incorporated in the model by assigning prior distributions to model inputs. A uniform distribution is commonly used when there is little knowledge about an uncertain parameter in the model except that its value has to lie anywhere within fixed bounds. The risk attitude θ_{VaR} is sampled from the uniform distribution $\mathrm{U}(0.5\%, 50\%)$ reflecting a range of investment assumptions from extremely risk averse to risk neutral. In the GB power system with 60 GW peak demand, θ_{markup} is sampled from $\mathrm{U}(0, 25)$ [GW]. The chosen range of θ_{markup} implies that the price uplift function is used under system conditions with a fairly tight capacity margin (range $0-25\%$). For 2015 onward, $VOLL$ is assumed to sample from $\mathrm{U}(1000, 20000)$ [£/MWh] over which we wish to achieve the desired understanding through sensitivity analysis; this is a reasonable range according to the study in [35] where the VOLL is estimated for domestic, industrial and commercial electricity consumers in GB. A prior uniform distribution $\mathrm{U}(0.80, 1.20)$ is assigned to P_{carbon}. In the case that $P_{carbon} = 1.05$, the reference trend of carbon prices, which takes DECC's central forecast [36], would be shifted upwards by 5%. The uncertainty range of future carbon prices resulting from the chosen range of P_{carbon} is broadly in line with the range of DECC's carbon projections in [36]. The prior belief for the forcing input P_{gas} is a normal distribution $\mathrm{N}(1, 0.06^2)$, which indicates bias over the reference gas price level, and results in a range consistent with that estimated by DECC [26].

4.2. Computational time

The simulator was run in the Matlab/Simulink R2012a environment using an Intel(R) Core(TM) i5 − 3470 3.20GHz processor with 8.00GB RAM. The run time for a single simulation of the 30-year generation planning varied between 140 and 600 minutes, with 7 stochastically simulated years and 100 MC simulations for each investment decision. The GPM/SA code package that was developed by Los Alamos National Laboratory as described in [37] has been adapted in this paper and used for emulator development and valuations. In comparison with simulator runs, one emulator evaluation in the same environment took approximately 10^{-4} seconds, a speed ratio in the order of $10^7 \sim 10^9$. The extra time needed for developing the emulator is about 6 minutes excluding

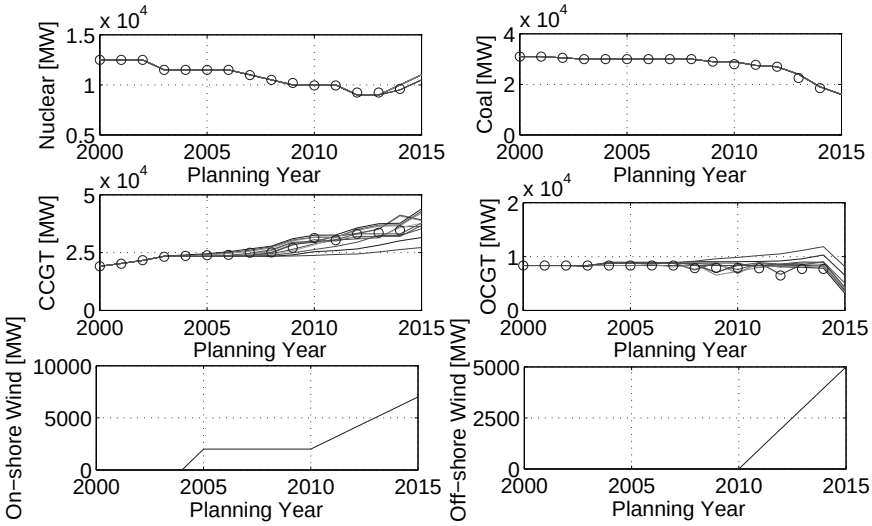

Figure 5: Simulated historical installed capacities of each thermal generation type given different design points compared with and observations

the time required for a handful of simulator runs (depending on the amount of training data required).

Traditional MC-based probabilistic SA that is directly applied to the simulator would take several months or even over one year for a thousand of simulation runs. However, with the developed emulator, the sensitivity task can be achieved within several seconds. The advantage of the emulator-based approach in saving computational time is clearly seen.

4.3. Calibration

It will be shown here how the simulator can be calibrated against historical observations of model output. In practice a good simulator is needed in that it is able to reasonably well reproduce the dynamics of observations (at least for some parameter values).

An emulator is built using 12 training data that are composed of 12 design points over the input space $(\theta_{VaR}, \theta_{markup})$ and the corresponding 12 scenarios of annual outcome of total thermal capacity in operation over the planning horizon $\mathcal{P}$. The observation data y_{obs}, against which the simulator is calibrated, consists of a single time series of $N = 12$ observations. We assume that any observation error is negligible. Fig. 5 gives a breakdown of simulated (in lines) and observed (in circles) installed thermal generation capacities at all design points, and also historically observed on-shore and off-shore wind capacities.

4.3.1. Validation of the emulator

Apart from the training data, 6 additional model runs on a maximin Latin hypercube design are used for validation. Fig. 6 shows that almost all the results

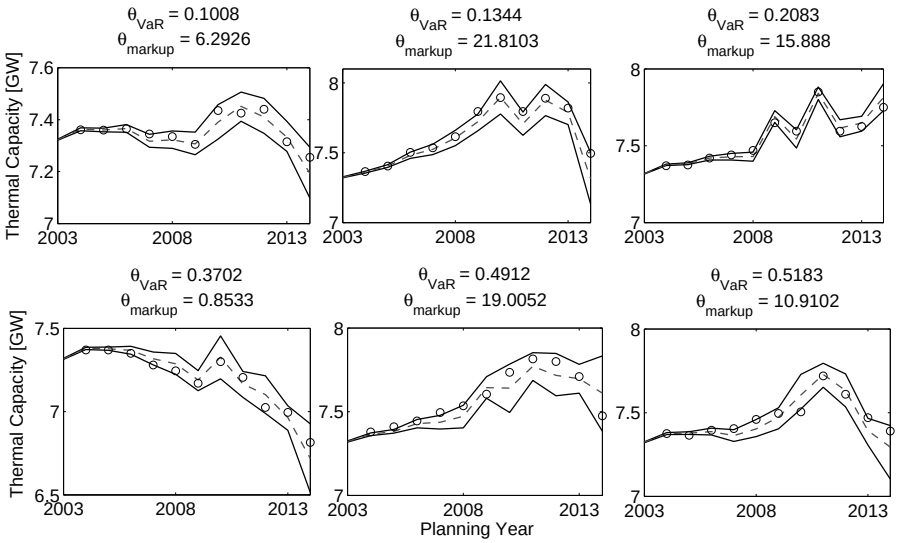

Figure 6: Predictions at data points for validation; Circles show the observation simulated by the simulator; Dashed and solid black lines indicate the mean and the the 5th and 95th percentiles of model output predicted by the emulator.

produced by the simulator are located within the 95% confidence intervals of model output predicted by the emulator, suggesting that the emulator performs well. Fig. 7 shows a boxplot of residuals (*i.e.*, differences between predicted values from the emulator and observed values from the simulator) at each planning year. The root mean-square error (RMSE) between the emulator's mean prediction and the simulator output is 46.4 MW. The ratio of RMSE to the mean value of the simulated output is 0.0062; this is small and indicates a good mean prediction by the emulator.

It is worth noting that a probabilistic prediction of generation projections at a point estimate of model inputs is estimated by the emulator here, whereas in [6, 18, 27] one path within the uncertainty window of the calibrated uncertain range would be seen, resulting in overconfident predictions. This is also true for future projections in Section 4.4. Since a full Bayesian approach has been taken in this paper, the model inputs, the model itself, and the model outputs of interest are all treated as uncertain and described through probability distributions.

The following calibration work is based on the validated emulator, with the aim of obtaining the posterior distributions of the calibration parameters, and also inferring the model discrepancy, so that these may be used to make future projections. This approach improves on validation work done in [27], where the simulation results at the assumed 'good' values of calibration parameters are graphically compared against observations, and issues such as model structure discrepancy are not addressed.

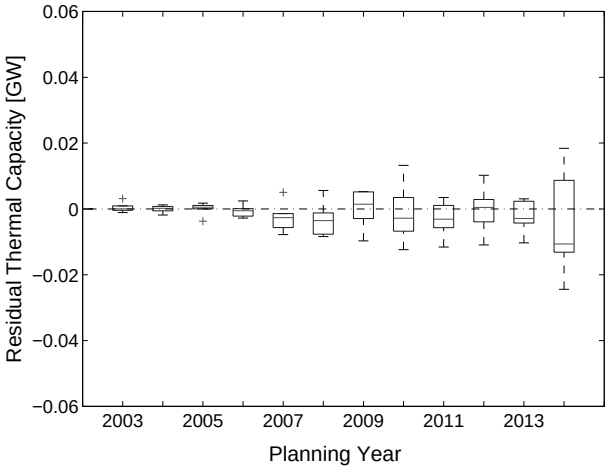

Figure 7: Boxplots of holdout residuals at each planning year

4.3.2. Parameter Calibration

Fig. 8 shows probability density functions for the marginal and bivariate posterior distributions of the two calibration parameters on the normalized $[0, 1]$ scale. As compared with their prior uniform distributions specified in Section 4.1, the posterior distributions are constrained in these two dimensions by removing input values that result in implausible outputs.

The posterior distribution of θ_{VaR} shows that the investor tends to be risk-averse, but it might be inappropriate to assume a very risk-averse investor given the probabilistic descriptions of fuel prices and demand growth rates in the simulator. The posterior distribution of θ_{markup} indicates that it is plausible for generators to receive an uplift payment when the capacity margin falls into the range $[4 - 23]$ GW (on the original scale).

4.3.3. Calibrated and Discrepancy-adjusted Simulator

The risk of overfitting using the calibrated emulator is relatively low compared with the simulator used in [27], due to the use of a posterior distribution covering a range of possible values for calibration parameters which is the case in our study (see Fig. 8) rather than a single combination of values. However, there is potential for overfitting if too many Normal kernels d_i are specified to model the discrepancy term δ, resulting in too many parameters that need calibrating given the limited number of observations available. The model discrepancy is expected to have a strong time persistence so that the normal kernels of δ are wide in the t direction. Here, the model discrepancy is represented by a restricted number of 3 weighted Normal kernels centered on years $(2005, 2009, 2013)$, each with a standard deviation of 2. Note that the observations at years $2001 - 2002$ are omitted from calibration because the investment decisions do not take effect

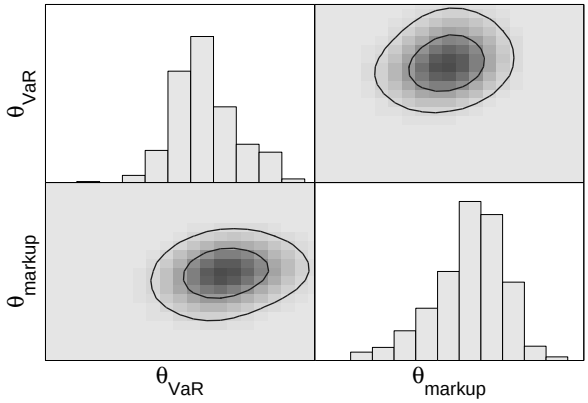

Figure 8: Marginal and bivariate posterior distributions of θ on the [0,1] scale.

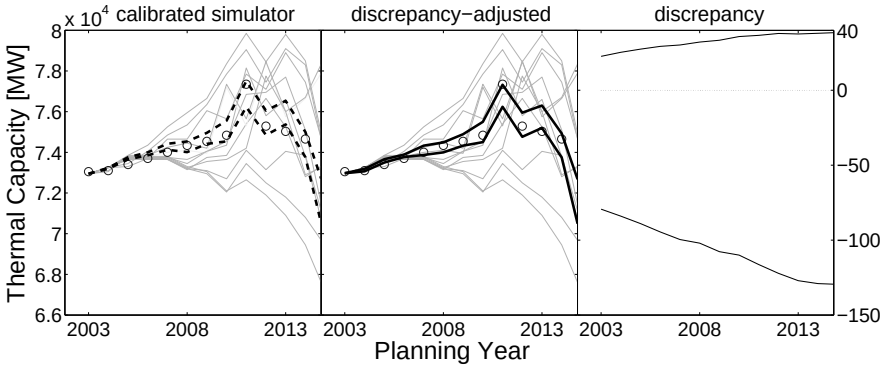

Figure 9: Circles show the historical data; In the left and center columns, dashed and solid black lines indicate a plausible range of installed thermal capacities, and grey lines correspond to the 12 simulated scenarios of thermal capacities during years $2003-2014$. Left: Calibrated simulations; Center: Discrepancy-adjusted calibrated simulations after adding the discrepancy term to the calibrated simulations; Right: Plausible range of the discrepancy term between emulated and observed values.

until 2003 due to construction delays associated with thermal power plants.

The input uncertainty (*i.e.*, the uncertainty in the calibration parameters), represented by the prior or posterior distributions, propagates into the output uncertainty, resulting in a probability distribution over the outputs. Fig. 9 shows how the calibration and the discrepancy term reduce the plausible output space when observations are available. A credible interval of $[5\%, 95\%]$ is taken as the plausible range in this paper. Without calibration, a wide range of simulator outputs is observed, as shown by the light grey lines in both the left and the centered graphs. With a calibrated simulator, a much narrower plausible range of simulator outputs is identified by the dashed black lines in the left column. After adding the discrepancy term to the calibrated simulations, the predicted range of simulator outputs, as quantified by the black lines in the centered graph, more closely matches the observation data compared with that predicted by the calibrated simulator. The plots in the right column quantify the plausible range of the model discrepancy, which has a much lower order than the model output and grows slightly over time.

4.4. Future projections using the discrepancy-adjusted and calibrated emulator

The aim of this section is to use the posterior distributions obtained from the calibration of the simulator against historical observations to obtain plausible future projections from the simulator. Additional uncertainty on the inputs (u, ω) takes effect in making future projections. Following a Bayesian approach allows for the combination of the simulator runs, the posterior distributions of the calibration parameters θ, and the model discrepancy.

As for calibration, the first stage for obtaining plausible future projections is building an emulator; this is trained using a group of 25 design points over the 5-dimensional input space $x = \{u, \theta, \omega\}$ and the corresponding 25 scenarios of annual installed thermal capacity over period $\mathcal{F}$. The same validation approach described in Section 4.3.1 was employed and similar results were obtained suggesting a good fit. All these design and validation points are sampled using a Latin-hypercube design over their prior ranges. Fig. 10 shows a breakdown of simulated scenarios of installed thermal capacities given a projection of future on-shore and off-shore wind capacity.

The validated emulator is used for making predictions in the second stage, where the control variables are fixed but the calibration and forcing parameters remain uncertain. In this case, the output uncertainty results from the uncertainty of forcing and calibration parameters as well as the approximation error of the emulator. Fig. 11 shows a probabilistic prediction of future thermal capacities at fixed values of $VOLL := 10000$ and $P_{carbon} := 1$. The grey lines in the left and center column show all the simulations obtained from the simulator. A comparison is made between the plausible range of thermal capacities predicted by the simulator before calibration (in dashed lines in the left column), and the calibrated and discrepancy-adjusted simulator (in black lines in the center column). As expected, a much wider plausible range of future thermal capacities is predicted by the simulator before calibration; in this case, the prior distributions of calibration parameters are used instead of their

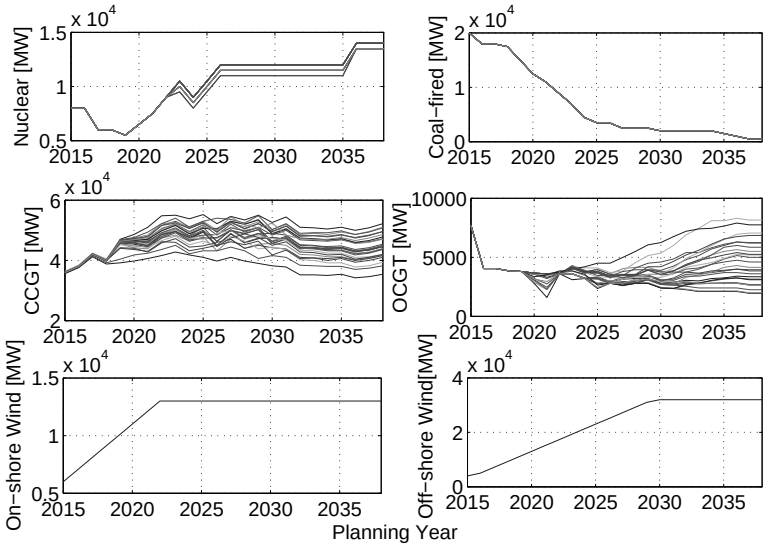

Figure 10: A breakdown of simulated generation projections across all the design points

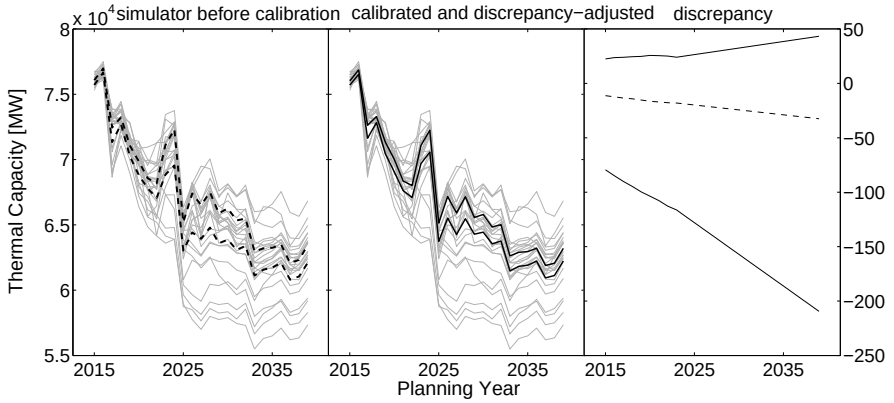

Figure 11: In the three columns, dashed and solid black lines indicate the plausible range of projected thermal capacities, and grey lines correspond to the simulated scenarios of thermal capacities over the planning horizon across the full range of design points. Left: Simulator before calibration; Center: Discrepancy-adjusted calibrated simulations; Right: Discrepancy term applied to future projections.

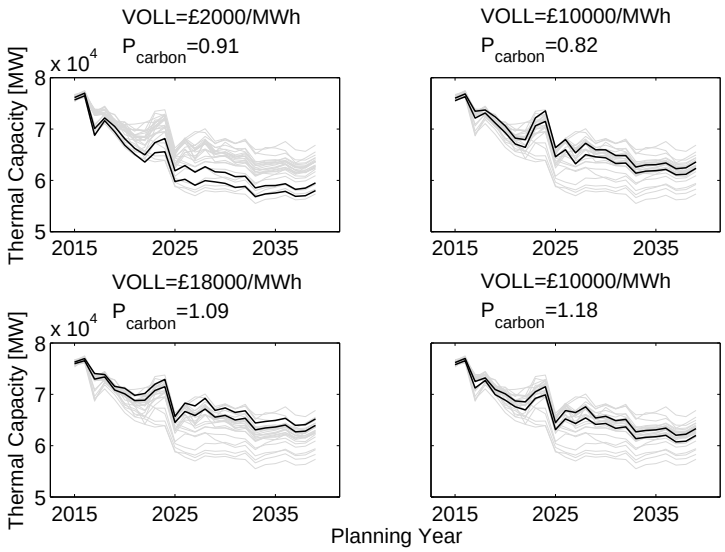

Figure 12: Predictions of thermal capacities at different values of $(VOLL, P_{carbon})$. Grey lines are simulation output and black lines indicate the plausible range of projected thermal capacities.

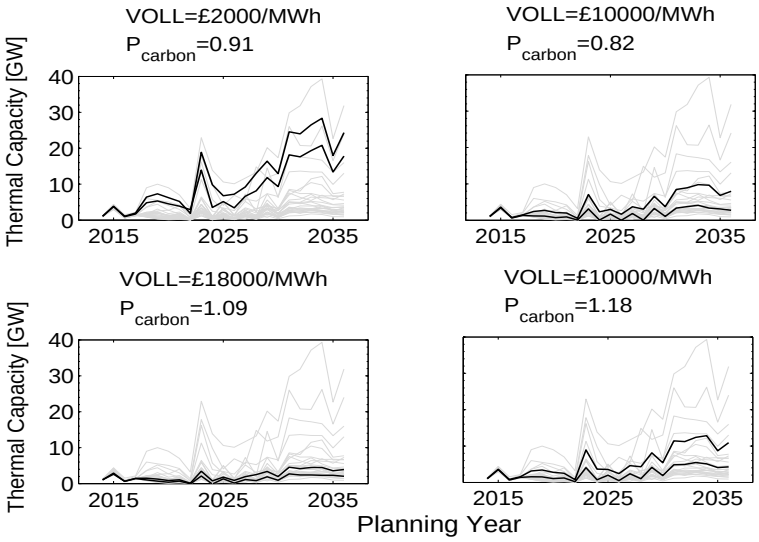

Figure 13: Predictions of LOLE profiles at different values of $(VOLL, P_{carbon})$. Grey lines are simulation output and black lines indicate the plausible range of the projected LOLE.

posterior distributions and no model discrepancy is accounted for. The right column in Fig. 11 shows the model discrepancy applied to future projections, which is consistent with the discrepancy term inferred from history matching for the first 12 years of simulation. In order to reflect the increasing uncertainty far into the future, the model discrepancy is assumed to increase by 5% of the average systematic error at every year from 2026 onwards. Although there is no guarantee that the modelled historical discrepancy would accurately apply to future projections, the discrepancy term added to the calibrated predictions can mitigate the risk of making overconfident projections.

To help understand the combined effects of $(VOLL, P_{carbon})$, the probabilistic predictions (*i.e.*, the uncertainty ranges) of thermal capacities and LOLE are presented in Fig. 12 and Fig. 13 respectively, at selected combinations of "high", "middle" and "low" values of both variables. LOLE, one of the outputs of the LTGI model and calculated in (1c), indicates the level of security of supply risk. In comparison with the GB standard of 3 hours per year LOLE [38], the risk of security of supply from year 2023 onwards for some choices of input settings can be very high, as shown by the grey lines in Fig. 13. The LOLE profiles are emulated using the Bayesian approach as described in Section 3.1. The right-hand graphs in Fig. 12 and Fig. 13 show that thermal capacities decline and the LOLE increases as the value of P_{carbon} increases (*i.e.*, the trend level of carbon prices increases) but the effect is small. One of the advantages of the probabilistic predictions is that it is natural and computationally efficient to determine the combination of $(VOLL, P_{carbon})$ with a high probability of keeping the $LOLE$ at each planning year below some set threshold.

4.5. Sensitivity Results

Sensitivity analysis is firstly conducted to study the individual and combined effects of calibration parameters on the installed thermal capacities for the emulator used for calibration. Fig. 14 shows variations in the output by varying one parameter while averaging over the other (instead of holding the other constant). Table 3 provides the variance contributions of each calibration parameter to the total variance of the model output in terms of the main effect index defined in (10). It is clear that the installed thermal capacity is much more sensitive to the change in the price markup model parameter θ_{markup} than the risk-preference parameter θ_{VaR}.

Table 3: Measures of sensitivity on the calibration parameters in history matching

Input variables	θ_{VaR}	θ_{markup}	$(\theta_{VaR}, \theta_{markup})$
% of variance	6	85	9

A more comprehensive SA is conducted for future projections under the input uncertainty. Table 4 and Table 5 provide the variance contributions of individual model inputs and those of two-input interactions to the overall variance associated with the projections of thermal capacity, respectively. It is clearly

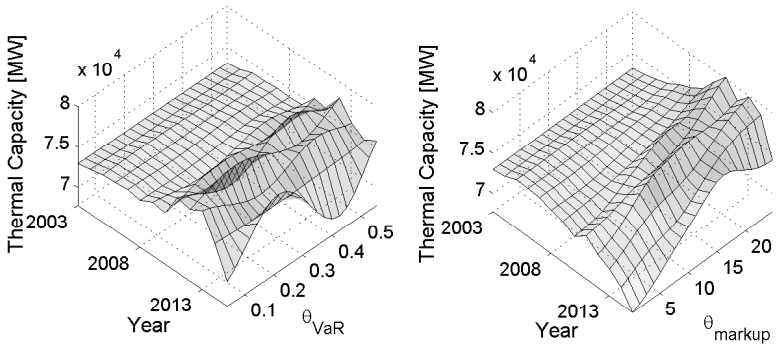

Figure 14: Sensitivity analysis of $\theta = \{\theta_{VaR}, \theta_{markup}\}$ in history matching

seen that the most important input parameters are $VOLL$ and θ_{markup}, as varying these has the largest effect on the output. An index of 57.6% indicates that uncertainty about $VOLL$ accounts for over half of the the overall uncertainty in the output. The results from SA highlight the significant long-term effect of scarcity pricing on incentivizing the investment of conventional generation capacity in future electricity markets with a high penetration of wind power.

Table 4: Sensitivity of the five input parameters in future projections

Input variables	$VOLL$	P_{carbon}	P_{gas}	θ_{VaR}	θ_{markup}
% of variance	57.6	2.3	0.8	0.6	34.9

Table 5: Sensitivity of two-input interactions in future projections

Inputs	$(VOLL, P_{carbon})$	$(VOLL, P_{gas})$	$(VOLL, \theta_{VaR})$
% of variance	0.46	0.13	0.01
Inputs	$(VOLL, \theta_{markup})$	(P_{carbon}, P_{gas})	$(P_{carbon}, \theta_{VaR})$
% of variance	1.84	0.12	0.02
Inputs	$(P_{carbon}, \theta_{markup})$	(P_{gas}, θ_{VaR})	$(P_{gas}, \theta_{markup})$
% of variance	0.53	0.02	0.33
Inputs	$(\theta_{VaR}, \theta_{markup})$	% of variance	0.13

4.6. Generality of the Bayesian framework

The case study in this Section has provided a full exemplar of how the Bayesian framework can be applied in a large, complex problem, showing all

important tasks of calibration, validation, uncertainty analysis and sensitivity analysis. As mentioned in Section 2, the Bayesian framework is very useful for those who rely on models to understand complex process involving uncertainty, and those who wish to know how much they can trust the model outputs. The unique characteristics found in different models mean that the emulator must be tuned for each particular model, e.g. the dimensionality of the input and the output space, and prior distributions for inputs, the emulator and the discrepancy. Besides, different parametric forms can be chosen for the emulator and the model discrepancy to reflect the characteristics of a particular model.

Examples of power system models to which the emulation methods described in this paper could be applied include: models discussed in [39, 40], which are used in GB Electricity Market Reform; a model presented in [13] to study the capacity market proposal of PJM; and models described in [7, 9–12] which are used for academic and industrial studies of electricity markets. The particular modeling methods used in these studies are different to those used in the LTGI model described in this paper. For example, methods for the modeling of uncertainties within the model (e.g., load, wind generation, fuel prices, policy options), energy dispatch (e.g., Dynamic Dispatch Model (DDM) [39], PLEXOS [40]) and investment logic may differ. A particular challenge for fitting an emulator to the DDM in [39] is the high-dimensional input space of this model, meaning that a careful selection of the inputs to be included in any emulator must be carried out. A key similarity between the DDM and the LTGI exemplar is that outputs for both models are in the form of a time-series, so the dimension reduction techniques and the form used for the model discrepancy described here could be applied directly to the DDM. The PJM capacity market model in [13] can be enhanced by employing the Bayesian emulation method to provide a more realistic and confident assessment of capacity market designs. Some model parameters (e.g., parameters in the utility function) that were originally chosen based on behavioral assumptions, can be more rigorously calibrated against historical observations of model outputs of interest by fitting an emulator to the model retaining the same main features but with an energy-only market design where historical data were observed.

5. Conclusion

This paper has demonstrated the use of a Bayesian method of emulation on a computationally expensive long-term generation investment model. This Bayesian method allows for calibrating uncertain model parameters, quantifying the model discrepancy, and making probabilistic predictions of the model output that are consistent with historical observations, as well as conducting SA in a formal and efficient manner. The emulator allows management of uncertainties arising from the limited number of evaluations of the LTGI model and its imperfect science. Such calibration and uncertainty analysis approaches are necessary in linking modeling results to the real system which the model is intended to represent, and hence enabling better decision making for both investors and policymakers.

In our GB case study, for model inputs, the uncertainty on regulatory decisions (e.g., VOLL and carbon prices), future gas prices, and calibration parameters (e.g., investor's risk preference and energy price markup) are considered and specified with prior distributions. For model outputs, both the real-world observations and the simulated outputs of installed thermal capacities are considered as well as simulated LOLE profiles. Future projections are obtained using a combination of the simulator, calibrated model parameters and model discrepancy. The calibration results and the predictions in the case study have been compared the validation performed by Eager *et al.* in [6, 18, 27], as discussed in Section 4.3.1. Sensitivity analysis results quantitatively show that the investment decisions are most sensitive to the two factors affecting scarcity pricing – value of lost load and price markup. Our uncertainty about system capacity adequacy can be significantly reduced by reducing policy risks.

Acknowledgements

The authors express their thanks to colleagues at Durham University for many valuable discussions during the project, and express particular thanks to Dan Eager for providing the code of his generation investment model. They also acknowledge discussions with Dan Eager, Janusz Bialek and Michael Goldstein. This work was supported by the Chinese Scholarship Council, and by the following EPSRC grants: EP/K03832X/1 and EP/K02115X/1.

References

[1] P. L. Joskow, Capacity payments in imperfect electricity markets: Need and design, Utilities Policy 16 (3) (2008) 159–170.

[2] G. Muratore, Incentive based energy market design, European Journal of Operational Research 213 (2) (2011) 422–429.

[3] A. Botterud, G. Doorman, Generation investment and capacity adequacy in electricity markets, International Association for Energy Economics.

[4] R. Gross, W. Blyth, P. Heptonstall, Risks, revenues and investment in electricity generation: Why policy needs to look beyond costs, Energy Economics 32 (4) (2010) 796 – 804, policymaking Benefits and Limitations from Using Financial Methods and Modelling in Electricity Markets.

[5] W. Steggals, R. Gross, P. Heptonstall, Winds of change: How high wind penetrations will affect investment incentives in the {GB} electricity sector, Energy Policy 39 (3) (2011) 1389 – 1396.

[6] D. Eager, B. Hobbs, J. Bialek, Dynamic modeling of thermal generation capacity investment: Application to markets with high wind penetration, Power Systems, IEEE Transactions on 27 (4) (2012) 2127–2137.

[7] M. Assili, M. H. D.B., R. Ghazi, An improved mechanism for capacity payment based on system dynamics modeling for investment planning in competitive electricity environment, Energy Policy 36 (10) (2008) 3703–3713.

[8] S. Arango, E. Larsen, Cycles in deregulated electricity markets: Empirical evidence from two decades, Energy Policy (5) (2011) 2457–2466.

[9] F. Olsina, F. Garcés, H.-J. Haubrich, Modeling long-term dynamics of electricity markets, Energy Policy 34 (12) (2006) 1411–1433.

[10] L. de Vries, P. Heijnen, The impact of electricity market design upon investment under uncertainty: The effectiveness of capacity mechanisms, Utilities Policy 16 (3) (2008) 215–227.

[11] M. Hasani, S. H. Hosseini, Dynamic assessment of capacity investment in electricity market considering complementary capacity mechanisms, Energy 36 (1) (2011) 277–293.

[12] G. Doorman, A. Botterud, Analysis of generation investment under different market designs, Power Systems, IEEE Transactions on 23 (3) (2008) 859–867.

[13] B. Hobbs, M.-C. Hu, J. Inon, S. Stoft, M. Bhavaraju, A dynamic analysis of a demand curve-based capacity market proposal: The pjm reliability pricing model, Power Systems, IEEE Transactions on 22 (1) (2007) 3–14.

[14] A. J. Pereira, J. T. Saraiva, Generation expansion planning (GEP) – a long-term approach using system dynamics and genetic algorithms (GAs), Energy 36 (8) (2011) 5180–5199.

[15] C.-K. Han, D. Hur, J.-M. Sohn, J.-K. Park, Assessing the impacts of capacity mechanisms on generation adequacy with dynamic simulations, Power Systems, IEEE Transactions on 26 (4) (2011) 1788–1797.

[16] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index, Computer Physics Communications 181 (2) (2010) 259 – 270.

[17] A. Borison, Electric power resource planning under uncertainty: Critical review and best practices, Online: www.thinkbrg.com/media/publication/ (2014).

[18] D. Eager, Dynamic modelling of generation capacity investment in electricity markets with high wind penetration, Online: http://hdl.handle.net/1842/6264 (2012).

[19] B. Palmintier, M. Webster, Impact of unit commitment constraints on generation expansion planning with renewables (2011) 1–7.

[20] I. Herrero, P. Rodilla, C. Batlle, Electricity market-clearing prices and investment incentives: The role of pricing rules, Energy Economics 47 (0) (2015) 42 – 51.

[21] A. O'Hagan, Bayesian analysis of computer code outputs: A tutorial, Reliability Engineering & System Safety 91 (10C11) (2006) 1290–1300.

[22] S. Conti, A. O'Hagan, Bayesian emulation of complex multi-output and dynamic computer models, Journal of Statistical Planning and Inference 140 (3) (2010) 640 – 651.

[23] J. E. Oakley, A. O'Hagan, Probabilistic sensitivity analysis of complex models: A Bayesian approach, Journal of the Royal Statistical Society, Series B 66 (2002) 751–769.

[24] D. Higdon, J. Gattiker, B. Williams, M. Rightley, Computer model calibration using high-dimensional output, Journal of the American Statistical Association 103 (482) (2008) 570–583.

[25] R. S. Pindyck, The long-run evolution of energy prices, The Energy Journal 20 (2) (1999) 1–27.

[26] DECC fossil fuel price projections, Online: www.gov.uk/government/uploads/system/uploads/attachment_data/file/212521/130718_decc-fossil-fuel-price-projections.pdf (2013).

[27] D. Eager, J. Bialek, T. Johnson, Validation of a dynamic control model to simulate investment cycles in electricity generating capacity, in: Power and Energy Society General Meeting, 2010 IEEE, 2010, pp. 1–8.

[28] M. Morris, T. Mitchell, Exploratory designs for computer experiments, Journal of Statistical Planning and Inference 43 (1995) 381–402.

[29] D. Gamerman, H. F. Lopes, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd Edition, CRC Press, 2006, Ch. 7.

[30] L. S. Bastos, A. O'Hagan, Diagnostics for gaussian process emulators, Technometrics 51 (4) (2009) 425–438.

[31] H. Abdi, L. J. Williams, Principal component analysis, Wiley Interdisciplinary Reviews: Computational Statistics 2 (4) (2010) 433–459.

[32] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464.

[33] D. Higdon, Space and space-time modeling using process convolutions (2002) 37–56.

[34] C. P. Robert, G. Casella, Monte Carlo statistical methods, Springer-Verlag Inc, 1999.

[35] L. Economics, The value of lost load (VoLL) for electricity in great britain, www.gov.uk/government/uploads/system/uploads/attachment_data/file/224028/value_lost_load_electricty_gb.pdf (2013).

[36] D. of Energy & Climate Change, Carbon values used in DECC's energy modelling, Online: www.gov.uk/government/uploads/system/uploads/attachment_data/file/48185/3138-carbon-values-decc-energy-modelling.pdf (2011).

[37] J. G. B. R. Kary Myers, Dave Higdon, A detailed example of using the gpm/sa code, Online: www.lanl.gov/org/padste/adtsc/computer-computational-statistical-sciences/_assets/docs/gpmsa-ball-drop-tutorial-jan-2012.pdf (2012).

[38] GB electricity capacity margin, www.raeng.org.uk/publications/reports/, filename:gb-electricity-capacity-margin (2013).

[39] DECC dynamic dispatch model (DDM), Online: www.gov.uk/government/uploads/system/uploads/attachment_data/file/65709/ (May 2012).

[40] Dynamics of GB generation investment, Online: www.redpointenergy.co.uk/files/final_report_final.pdf (2006).